

1. 上课须知
2. 上次内容总结
3. 上次作业
4. 本次内容预告
  - 4.1. Hive的语法支持总结和执行原理总结
  - 4.2. Hive的groupby, join, distinct等语法实现原理总结
  - 4.3. Hive的SQL语句转换成MapReduce的机制详细解读
  - 4.4. Hive源码级全流程大图详解
  - 4.5. 阅读源码的特别方法 (IDEA)
  - 4.6. 关于HQL执行的结果解析
  - 4.7. 源码阅读的核心经验技巧分享
5. 总结
6. 作业

## 1. 上课须知

---

课程主题: Hive 的 SQL 编译成 MapReduce 源码剖析 (源码分析)

上课时间: 20:00 - 23:00

课件休息: 21:30 左右 休息10分钟

课前签到: 如果能听见音乐, 能看到画面, 请在直播间扣 666 签到

## 2. 上次内容总结

---

史诗级宇宙最全30条Hive性能调优全详解

- 1、深度剖析Hive架构设计和工作原理
- 2、史诗级Hive性能调优30条最佳实践
- 3、超经典Hive调优企业级案例实践

## 3. 上次作业

---

见文档: Hive最终大作业.pdf

## 4. 本次内容预告

---

- 1、Hive的语法支持总结和执行原理总结
- 2、Hive的groupby, join, distinct等语法实现原理总结
- 3、Hive的SQL语句转换成MapReduce的机制详细解读
- 4、Hive源码级全流程大图详解
- 5、源码阅读的核心经验技巧分享

## 4.1. Hive的语法支持总结和执行原理总结

见文档：Hive的HQL的编译原理

## 4.2. Hive的groupby, join, distinct等语法实现原理总结

select ... distinct()... from ... join ... on ... where .... group by .... having .... order by ... limit .....  
union.... in/exists ..... case... when...

见文档：Hive的HQL的编译原理

## 4.3. Hive的SQL语句转换成MapReduce的机制详细解读

见文档：Hive的HQL的编译原理

## 4.4. Hive源码级全流程大图详解

先按照图解，预习一下 Hive 的 SQL 到底是怎么解析成 MapReduce 的：

要高清大图去这儿：<https://blog.csdn.net/zhongqi2513/article/details/107153698>

## 4.5. 阅读源码的特别方法（IDEA）

查看类的结构：alt + 1

查看类的继承结构：ctrl + alt + U

过去使用的类的历史：ctrl + E

查看类的子类：ctrl + H

## 4.6. 关于HQL执行的结果解析

```
hive (studentdb)> select * from student limit 3;
OK
student.id      student.name    student.sex     student.age
student.department
2030-08-08 21:39:57,059 INFO [ba0ab48a-366f-4199-9076-8a33ee608593 main]
mapred.FileInputFormat: Total input paths to process : 1
95002  刘晨曦  女      19      IS
95017  王凤娟  女      18      IS
95018  王一伟  女      19      IS
Time taken: 0.298 seconds, Fetched: 3 row(s)
```

标准输入: select \* from student limit 3;

标准输出: 后面的所有, 分为四个组成

```
1、OK
2、student.id      student.name    student.sex     student.age
student.department
3、正常SQL结果
95002  刘晨曦  女      19      IS
95017  王凤娟  女      18      IS
95018  王一伟  女      19      IS
4、Time taken: 0.298 seconds, Fetched: 3 row(s)
```

休息10分钟, 21:44 继续

## 4.7. 源码阅读的核心经验技巧分享

- 1、了解大概原理
  - 大致的启动过程
  - 大致的心跳机制流程
  - 大致的任务提交过程
  - hive的SQL编译: 大概是怎样的过程
- 2、场景驱动
  - hdfs dfs -put /a /b
  - 启动流程? start-dfs.sh ---> NameNode DataNode
  - 数据读写流程? hadoop fs -put a.txt /
  - 任务执行机制? hadoop jar xxxx.jar com.nx.wc.WordCount
- 3、找入口
  - 启动集群的命令 start-dfs.sh
  - hadoop-daemon.sh start namenode
  - NameNode.main()
  - hadoop-daemon.sh start datanode
  - 提交任务的命令
- 4、理主线
  - namenode启动httpserver
  - namenode加载元数据
  - namenode启动RPCserver
- 5、看源码注释
  - 类注释 NameNode Lock
  - 成员变量注释
  - 成员方法注释

## 6、代码结构：方法的逻辑结构

### 1、参数解析 和 权限控制，总之为核心业务做准备

这种类似于跟核心业务逻辑无关的所谓初始化动作，都在方法的最开始

`if else ...` 做各种判断 大概率是无意义的

### 2、`try catch`， 一般来说，核心方法，都藏于`try`中，异常以及容错处理，都藏于 `catch` 中

核心逻辑代码一般都在 `try` 的最后。

如果`catch`中写了大量的代码：处理异常 + 容错方案

`finally: close stop .....`

### 3、状态处理， 一般用来处理核心业务的结果数据

如果 `finally` 后面还有代码：

核心逻辑分好几步，每个步，都有一句核心逻辑代码 每一句核心代码做单独的`try`控制

最后的状态的处理

### 4、收尾，回收资源相关

## 7、作图

啥都没有图好使！自从画了图，一辈子忘不了。

# 5. 总结

## 1、Hive的语法支持总结和执行原理总结

## 2、Hive的`groupby`, `join`, `distinct`等语法实现原理总结

## 3、Hive的SQL语句转换成MapReduce的机制详细解读

## 4、Hive源码级全流程大图详解

## 5、源码阅读的核心经验技巧分享

# 6. 作业

Hive 影评案例，见资料！10道题目！